اکوهیدرولوژی

دورهٔ ۴، شمارهٔ ۱، بهار ۱۳۹۶، ص ۳۷-۲۹

تعیین اثر پیش‌پردازش داده بر عملکرد شبکهٔ عصبی مصنوعی به‌منظور پیش‌بینی بارش ماهانه در شهرستان آباده

مهدی بهرامی[1*]، محمدجواد امیری[2]، فاطمه رضایی مهارلویی[3]، کرامت‌الله غفاری[4]

۱ و ۲. استادیار، گروه علوم و مهندسی آب، دانشکدهٔ کشاورزی، دانشگاه فسا

۳. دانشجوی کارشناسی ارشد آبیاری و زهکشی، گروه علوم و مهندسی آب، دانشکدهٔ کشاورزی، دانشگاه فسا

٤. مربی گروه فناوری اطلاعات، دانشکدهٔ مهندسی، دانشگاه فسا

(تاریخ دریافت ۱۳۹۵/۰۸/۲۰؛ تاریخ تصویب ۱۳۹۵/۱۰/۲۰)

چکیده

از آنجا که بسیاری از سری‌های زمانی نرمال نیستند، باید آنها را قبل از هرگونه تحلیل و مدل‌سازی نرمال کرد و این کار توسط توابع تبدیل صورت می‌گیرد. در پژوهش حاضر با استفاده از اطلاعات میانگین بارش ماهانه، کمترین و بیشترین دما و رطوبت ایستگاه سینوپتیک شهرستان آباده در بازهٔ زمانی ۱۳۵۵ تا ۱۳۹۲ به‌صورت نرمال‌شده و خام به‌عنوان ورودی‌های شبکهٔ پرسپترون چندلایه، بارش ماه آیندهٔ شهرستان پیش‌بینی شد. برای نرمال‌سازی داده‌های هواشناسی، پس از بررسی وجود داشتن یا نداشتن داده‌های گم‌شده و پرت از سه روش نرمال‌سازی مینیمم‌ـ ماکزیمم، رتبه‌ای و آمارهٔ نرمال استاندارد استفاده شد. پس از به‌دست‌آوردن بهترین ساختار شبکه با استفاده از آزمون و خطا برای هر روش از مقایسهٔ بهترین ساختارهای هر روش با یکدیگر، روش مینیمم‌ـ ماکزیمم با ساختار شبکهٔ سه‌لایه و تعداد ۱۳ نورون در لایهٔ پنهان با مقدار R=۰/۹۲ و MSE=۰/۱۲ در مقایسه با دیگر روش‌ها به‌عنوان بهترین روش انتخاب شد. نتایج آنالیز حساسیت انجام‌شده نیز نشان داد مدل به حذف پارامتر بیشترین رطوبت بیشتر از سایر پارامترها حساسیت نشان داد. پس از آن نیز دمای حداکثر بیشترین تأثیر را بر پیش‌بینی بارش داشت. همچنین مقایسهٔ عملکرد شبکه با تعداد ورودی‌های مختلف نشان داد شبکه با داشتن دو ورودی شامل کمترین دما و رطوبت با مقدار MSE =۰/۱۳ در مواقعی که کمبود داده وجود دارد نسبت به‌تعداد پنج ورودی به نتیجهٔ خوبی رسید.

کلیدواژگان: آمارهٔ نرمال استاندارد، بارندگی، مینیمم‌ـ ماکزیمم، نرمال‌سازی داده، نرمال‌سازی رتبه‌ای.

* نویسنده مسئول

Email: mehdibahrami121@gmail.com

Archive of SID

www.SID.ir

مقدمه

پایه و اساس مطالعات هیدرولوژی، داده‌های آماری مورد قبول است که از یک طرف موجب کوتاه‌ترشدن مدت مطالعات و از طرف دیگر، موجب برآورد دقیق‌تر پارامترهای هدف و کاهش هزینه‌های اجرایی و خسارت‌های بعدی ناشی از اجرای طرح‌ها می‌شود [۱]. با توجه به محدودیت‌هایی مانند نبود اطلاعات بارش در مقیاس‌های مکانی و زمانی مناسب و وابستگی آن به عوامل بسیاری از جمله دما، فشار، سرعت و جهت باد و همچنین غیرخطی‌بودن این فرایند عملاً استفاده از مدل‌هایی با مبنای فیزیکی مقدور نیست. امروزه به موازات مدل‌های متداول موجود، روش‌های جدیدتری نیز برای پیش‌بینی ابداع شده‌اند. شبکه‌های عصبی مصنوعی نوعی ابزار جدید هستند که در سیستم‌های غیرخطی و نامعین که روابط بین اجزا و پارامترهای سیستم به‌خوبی شناخته شده اند و توصیف‌پذیر نیستند، قادر به تحلیل و شبیه‌سازی‌اند [۲].

به‌دلیل خطاهای انسانی و ابزاری ممکن است مواردی مانند ثبت‌نشدن آمار، ثبت آمار غلط، خرابی یا ازبین‌رفتن دستگاه‌های اندازه‌گیری یا تشخیص داده‌های پرت و حذف آنها با عنوان داده‌های گم‌شده پیش آید، بنابراین تخمین و برآورد این داده‌ها برای استفاده در مدل‌ها ضروری است [۳]. باید قبل از به‌کارگیری این داده‌ها در مطالعات، نقایص آنها را برطرف و داده‌های گم‌شده را بازسازی کرد. بنابراین، دستیابی به روش‌های صحیح بازسازی داده‌ها ضروری به‌نظر می‌رسد. به همهٔ مراحلی که قبل از مدل‌سازی و به‌منظور آماده‌سازی داده‌ها برای کم‌کردن خطا صورت می‌گیرد، پیش‌پردازش گفته می‌شود. به‌طورخلاصه می‌توان گفت که پیش‌پردازش داده‌ها شامل همهٔ تبدیلاتی است که روی داده‌های خام انجام می‌شود و آنها را به‌صورتی در می‌آورد که برای پردازش‌های بعدی نظیر استفاده در دسته‌بندی، پیش‌بینی و تخمین ساده‌تر و مؤثرتر می‌سازد [۴]. از جمله فرایندهای پیش‌پردازش می‌توان حذف داده‌های گم‌شده، حذف داده‌های پرت و نرمال‌سازی داده‌ها را نام برد.

روش‌های متعددی برای تخمین داده‌های گم‌شده وجود دارد که بسته به شرایط هر ایستگاه ممکن است یک روش خاص بهترین نتیجه را در پی داشته باشد. معمولاً برای برآورد داده‌های گم‌شده در یک ایستگاه از ایستگاه‌های مجاور آن استفاده می‌شود که همبستگی بیشتری دارند [۳].

از طرفی، ممکن است ویژگی‌هایی در داده شامل مقادیری باشند که در محدوده یا دامنهٔ متفاوتی قرار داشته باشند. این ویژگی‌های با مقادیر بزرگ ممکن است اثر بسیار زیادتری در تابع هزینه نسبت به ویژگی‌های با مقادیر کوچک داشته باشند. این مشکل با نرمال‌کردن ویژگی‌ها برطرف خواهد شد به‌طوری که مقادیرشان در دامنه‌های مشابه قرار گیرد. به‌دلیل اینکه شرط اصلی و لازمهٔ استفاده از داده‌های هیدرولوژیکی در تجزیه و تحلیل‌های آماری و مدل‌سازی، نرمال‌بودن داده‌هاست، بنابراین قبل از استفاده از داده‌ها باید از نرمال‌بودن آنها مطمئن بود [۵].

ادب و همکارانش (۱۳۸۷) برای نرمال‌کردن داده‌های بارش سالانهٔ استان خراسان رضوی در تهیهٔ نقشه‌های بارش سالانه، از تبدیل باکس ـ کاکس استفاده کردند. نتایج نشان‌دهندهٔ مناسب‌بودن این تبدیل بود [۶].

حمیدی و امامقلی‌زاده (۱۳۸۸) برای نرمال‌سازی داده‌های سری زمانی دبی سالانهٔ رودخانهٔ مارون، از توابع لگاریتمی، جذر و نظیر آن استفاده کردند و در نهایت با توجه به مقدار ضریب چولگی، تبدیل لگاریتمی را برای نرمال‌ساختن داده‌های استفاده‌شده انتخاب کردند [۷].

شفیعی و همکارانش (۱۳۹۰) نیز برای نرمال‌کردن داده‌های استفاده‌شده در شبیه‌سازی تصادفی شدت خشکسالی براساس شاخص پالمر و ثقفیان و همکارانش (۱۳۹۰) برای نرمال‌کردن داده‌های بارش سالانهٔ استان فارس از تبدیل باکس ـ کاکس استفاده کردند [۸ و ۹].

احمدی و همکارانش (۱۳۹۱) برای نرمال‌کردن داده‌های دبی جریان ماهانه و روزانهٔ رودخانهٔ باراندوزچای از توابع مختلف مانند لگاریتم، جذر، تابع توانی و باکس ـ کاکس استفاده کردند و با توجه به ضریب چولگی نزدیک به صفر، تابع لگاریتم را برای نرمال‌کردن داده‌های منطقهٔ مطالعه‌شده، انتخاب کردند [۱۰].

نیک‌منش و طالب بیدختی (۱۳۹۱) دو مدل تئوری موجک سری‌های زمانی را با مدل‌سازی بارندگی ماهانهٔ ارسنجان استان فارس مقایسه کردند و برای نرمال‌کردن داده‌های بارندگی ماهانهٔ منطقهٔ مطالعه‌شده از تبدیل باکس ـ کاکس استفاده کردند [۱۱].

ناظری تهرودی و همکارانش (۱۳۹۱) با مقایسهٔ روش‌های رایج نرمال‌سازی برای مدل‌سازی دما با استفاده

Archive of SID

www.SID.ir

از مدل‌های پریودیک ARMA، تبدیل توانی را با توجه به مقدار ضریب چولگی انتخاب کردند [۱۲].

ناظری تهرودی و همکارانش (۱۳۹۲) برای نرمال‌سازی داده‌های ماهانهٔ دما و بارندگی ایستگاه سینوپتیک سنندج در مدل‌سازی اقلیم آیندهٔ کردستان، به‌ترتیب از تبدیل توانی و لگاریتم استفاده کردند و ضریب چولگی، این نتیجه را تأیید کرد [۱۳].

ناظری تهرودی و همکارانش (۱۳۹۲) در ارزیابی مدل‌های ARIMA و PARMA برای مدل‌سازی و پیش‌بینی بیشترین سرعت باد ایستگاه سینوپتیک بندرعباس از داده‌های سری زمانی ماهانهٔ ۵۵ سالهٔ منطقهٔ مطالعه‌شده استفاده کردند و برای نرمال‌کردن این سری زمانی از تبدیل‌های گاما و لگاریتم برحسب ماه‌های مختلف و ضریب چولگی بهره بردند [۱۴].

از آنجا که در بسیاری از منابع پیش‌پردازش و همچنین نرمال‌سازی داده‌ها که یکی از عوامل تأثیرگذار بر نتایج پیش‌بینی است به‌صورت اساسی در نظر گرفته نمی‌شود، بنابراین در پژوهش حاضر با استفاده از اطلاعات هواشناسی ایستگاه سینوپتیک شهرستان آباده واقع در استان فارس به‌صورت نرمال‌شده و خام به‌عنوان ورودی‌های شبکهٔ عصبی مصنوعی پرسپترون چندلایه، بارش ماه آیندهٔ شهرستان پیش‌بینی شد و در نهایت بهترین ساختار شبکه و بهترین روش نرمال‌سازی انتخاب شد.

مواد و روش‌ها

در این پژوهش با استفاده از اطلاعات هواشناسی ایستگاه سینوپتیک شهرستان آباده واقع در استان فارس شامل میانگین بارش ماهانه (میلی‌متر)، کمترین و بیشترین دمای ماهانه (سانتی‌گراد) و کمترین و بیشترین رطوبت ماهانه (درصد) در بازهٔ زمانی ۱۳۵۵ تا ۱۳۹۲بارش ماه آینده توسط شبکهٔ عصبی مصنوعی پرسپترون چندلایه (MLP[1]) مدل‌سازی شد [۱۵ـ ۱۷] زیرا شبکه‌های عصبی مصنوعی ابزار جدیدی هستند که در سیستم‌های غیرخطی و نامعین که روابط بین اجزا و پارامترهای سیستم به‌خوبی شناخته‌شده و توصیف‌پذیر نیستند، قادر به تحلیل و شبیه‌سازی‌اند. قبل از مدل‌سازی و تخمین بارش به پیش‌پردازش داده‌های موجود پرداخته شد. بدین‌منظور داده‌های گم‌شده و داده‌های پرت مشخص و بازسازی شدند و سه روش نرمال‌سازی داده‌ها شامل مینیمم‌ـ ماکزیمم، نرمال استاندارد و نرمال‌سازی رتبه‌ای برای پیش‌پردازش داده‌ها استفاده شد. به‌منظور بررسی اثر پیش‌پردازش، داده‌ها به دو صورت خام و نرمال‌شده به‌عنوان ورودی‌های شبکهٔ پرسپترون چندلایه استفاده و بارش ماه آیندهٔ شهرستان پیش‌بینی شد و خطای شبکه (MSE) و ضریب همبستگی (R) در همهٔ حالات بررسی و مقایسه شد و در نهایت بهترین ساختار شبکه و بهترین روش نرمال‌سازی انتخاب شد. برای پیش‌پردازش داده‌ها و نیز مدل‌سازی بارش توسط شبکه‌های عصبی مصنوعی از نرم‌افزار Matlab استفاده شد.

منطقهٔ مطالعه‌شده

ایستگاه سینوپتیک شهرستان آباده در شمال استان فارس با مختصات ۵۲ درجه و ۳۱ دقیقهٔ طول جغرافیایی، ۲۹ درجه و ۳۶ دقیقهٔ عرض جغرافیایی و ارتفاع ۲۰۳۰ متر از سطح دریا واقع شده است (شکل ۱).

شبکه‌های عصبی مصنوعی

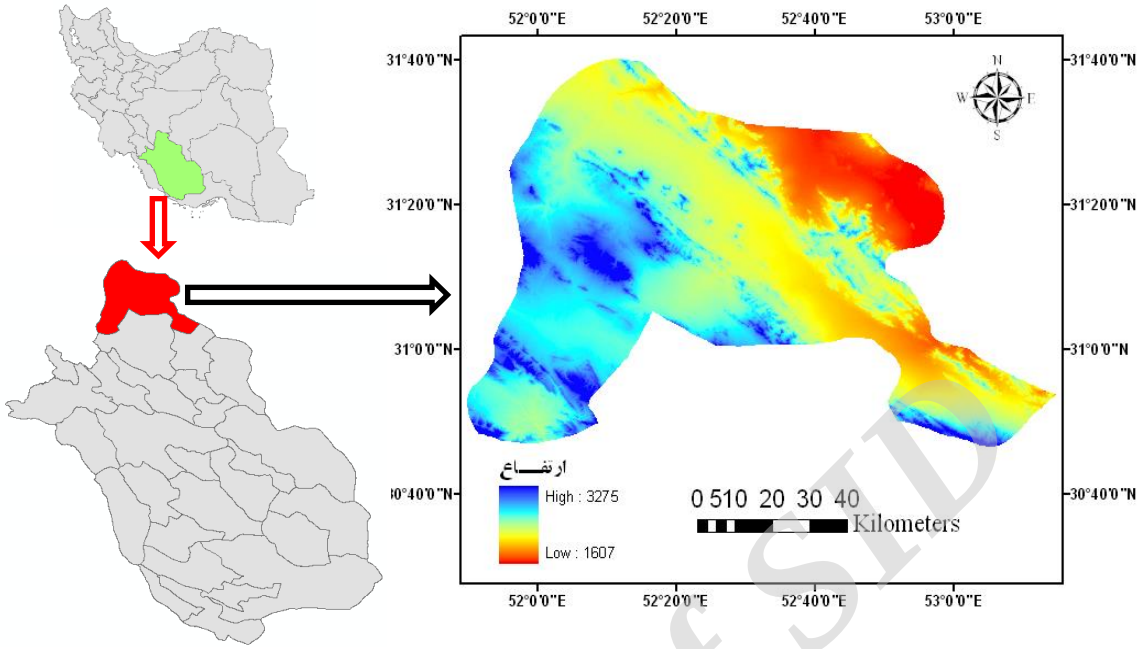
شبکهٔ عصبی مصنوعی به‌طور گسترده در تحقیقات هوش مصنوعی و در جایی کاربرد دارد که یک تابع تقریبی برای نگاشت غیرخطی بین پارامترهای ورودی و خروجی نیاز است [۱۸]. نخستین کاربرد عملی شبکه‌های عصبی مصنوعی توسط روزنبلات با معرفی شبکه‌های پرسپترون چندلایهٔ MLP انجام شد. به‌طور کلی، ساختار شبکهٔ عصبی MLP از یک لایهٔ ورودی، تعدادی لایهٔ پنهان و یک لایهٔ خروجی تشکیل شده است که برای معماری شبکه در هر لایه تعدادی نورون در نظر گرفته می‌شود.

در شبکه‌های عصبی وزن‌های اولیه، انتخاب تابع فعال‌سازی و انتخاب تعداد نورون‌های استفاده‌شده در لایهٔ میانی (پنهان) و همچنین تعداد این لایه‌ها به‌عنوان فاکتورهای یادگیری در نظر گرفته می‌شوند. شناخت و انتخاب صحیح این فاکتورها مهم است، زیرا نه‌تنها بر همگرایی شبکه تأثیر دارد، بلکه بر دقت پیش‌بینی و تخمین نیز اثر می‌گذارد [۱۸].

1. Multi-Layer Perceptron

Archive of SID

www.SID.ir



شکل ۱. موقعیت شهرستان آباده در استان فارس

طراحی یک شبکه عبارت است از تعیین تعداد لایه‌های مناسب، تعداد مناسب نورون‌ها در هر لایه و نوع تابع تحریک در هر لایه به‌گونه‌ای که خطای آموزش و آزمایش شبکه به حداقل برسد [۱۹]. برای به‌دست‌آوردن تعداد لایهٔ میانی و همچنین تعداد گره مناسب، می‌توان از روش لونبرگ‌ـ مارکوارت[1] استفاده کرد که از سریع‌ترین روش‌های آموزش است [۲۰].

در این پژوهش برای مدل‌سازی بارش با استفاده از شبکهٔ MLP از تعداد لایه‌های پنهان مختلف با تعداد نورون‌های متفاوت و توابع محرک تانژانت سیگموئید استفاده شد. همچنین از میان روش‌های مختلف آموزش، روش پس‌انتشار خطا با الگوریتم لونبرگ‌ـ مارکوارت به‌دلیل همگرایی سریع‌تر در آموزش شبکه استفاده شد.

روش‌های مختلف نرمال‌سازی داده‌ها

توزیع نرمال از متداول‌ترین و مهم‌ترین توزیع‌های احتمالی پیوسته در نظریهٔ احتمالات است که توزیع گوسی نیز نامیده می‌شود [۵]. در این مطالعه از سه تبدیل نرمال‌ساز رایج در هیدرولوژی استفاده شد که آماره‌های این آزمون‌ها به شرح زیر است:

1. Levenberg-Marquardt

الف) تبدیل مینیمم‌ـ ماکزیمم

تبدیل خطی را روی دادهٔ اصلی انجام می‌دهد و ارتباط بین مقادیر دادهٔ اولیه را حفظ می‌کند. اگر یک مقدار ورودی در خارج از محدودهٔ اولیهٔ داده قرار گیرد، خطای خارج محدوده رخ خواهد داد [۲۱]. در این تبدیل، داده‌ها به مجموعهٔ جدیدی تبدیل می‌شوند که در آن همهٔ مقادیر بین صفر و یک هستند. بدین‌منظور می‌توان از رابطهٔ ۱ استفاده کرد:

$$Y_t = \frac{X_t - X_{\min}}{X_{\max} - X_{\min}} \qquad (۱)$$

که در آن Y_t مقدار نرمال‌شده، X_t مقدار هریک از داده‌ها، $X_{\min}$ کمترین مقدار داده‌ها و $X_{\max}$ بیشترین مقدار داده‌ها است.

ب) نرمال‌سازی رتبه‌ای

با توجه به نمونهٔ داده‌شده برای ویژگی یک مؤلفه برای همهٔ داده‌ها مانند $x_1, x_2, \ldots, x_n$ باید ابتدا رتبهٔ (مرتبه) آماری داده را یافت و سپس مقدار ویژگی داده‌ها را با رتبهٔ نرمال‌شده متناظر آن جایگزین کرد [۱۹]:

$$Y_t = \frac{x_1, x_2, \ldots, x_n^{rank(x_i)-1}}{n-1} \qquad (۲)$$

Archive of SID

www.SID.ir

که در آن X_i مقدار ویژگی برای داده iام است. این روش همهٔ مقادیر ویژگی را به‌طور یکنواخت در محدودهٔ صفر و یک می‌نویسد. زمانی که بیشتر از یک داده با همان مقدار ویژگی وجود داشته باشد، میانگین رتبه برای آن مقدار اختصاص داده می‌شود [۱۹].

ج) تبدیل نرمال استاندارد (z-score)

مقادیر داده‌ها براساس میانگین و انحراف استاندارد نرمال‌سازی می‌شوند. مقدار X_t توسط فرمول زیر به مقدار Y_t نرمال می‌شود:

$$Y_t = \frac{X_t - X_{mean}}{S} \qquad (۳)$$

که در آن X_t مقدار اولیه (تغییرنیافته)، X_{mean} میانگین داده‌ها و S انحراف معیار آنهاست. این روش نرمال‌سازی هنگامی مفید است که مینیمم و ماکزیمم واقعی داده‌ها ناشناخته است و یا داده‌های پرتی وجود دارند که نرمال‌سازی مینیمم- ماکزیمم را تحت تأثیر قرار می‌دهند. همچنین زمانی‌که پارامترهای جمعیت (μ و σ) ناشناخته باشند، از این روش نرمال‌سازی استفاده می‌شود [۱۴].

معیارهای ارزیابی

برای بررسی و آزمون اعتبار شبکه‌ها، با استفاده از دو آمارهٔ زیر، عملکرد آنها ارزیابی شد:

میانگین مربعات خطا[1] (MSE):

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(P_{stm} - P_{obs})^2 \qquad (۴)$$

که در آن، P_{stm}: مقدار بارندگی شبیه‌سازی‌شده توسط شبکهٔ عصبی مصنوعی، P_{obs}: مقدار بارندگی ماهانهٔ واقعی (اندازه‌گیری‌شده در هر ایستگاه)، n: تعداد مشاهدات است. این مقدار هر چه به صفر نزدیک‌تر باشد، نشان‌دهندهٔ عملکرد مناسب مدل است [۲۳].

ضریب همبستگی (R): همبستگی بین مقادیر بارش مشاهده‌شده و پیش‌بینی‌شده توسط مدل از رابطهٔ زیر محاسبه شد [۲۴]:

$$R = \frac{\sum_{i=1}^{i=n}\left[P_{obs} - \overline{P_{obs}}\right]\left[P_{stm} - \overline{P_{stm}}\right]}{\sqrt{\sum_{i=1}^{i=n}\left[P_{obs} - \overline{P_{obs}}\right]^2}\sqrt{\sum_{i=1}^{i=n}\left[P_{stm} - \overline{P_{stm}}\right]}} \qquad (۵)$$

که در آن $\overline{P_{obs}}$ میانگین بارش مشاهده‌شده و $\overline{P_{stm}}$ میانگین بارش پیش‌بینی‌شده است. هر چه مقدار معیار R به یک نزدیک‌تر باشد، عملکرد مدل بهتر است.

آنالیز حساسیت

همچنین به‌منظور مشخص‌شدن پارامتری که بیشترین تأثیر را بر بارش دارد از آنالیز حساسیت مدل با روش و ساختار شبکهٔ بهینه استفاده شد. به این‌صورت که هریک از پارامترها به‌طور جداگانه حذف و سپس عملکرد شبکه مشاهده شد. هریک از اجراهای شبکه که بیشترین مقدار خطا را داشته باشد پارامتر حذف‌شدهٔ آن به‌عنوان پارامتری که بیشترین تأثیر را بر بارش دارد انتخاب می‌شود [۴ و ۲۵].

نتایج و بحث

از اطلاعات بارش ماهانهٔ سال‌های ۱۳۵۵ تا ۱۳۹۲ در ایستگاه سینوپتیک شهرستان آبادهٔ استان فارس به‌منظور پیش‌بینی بارش ماهانه استفاده شد. داده‌های موجود طی ۳۷ سال به‌صورت سری زمانی مرتب شد. قبل از مدل‌سازی بارش با استفاده از شبکه‌های عصبی مصنوعی، تبدیلات نرمال‌ساز روی داده‌های موجود انجام شد که نتیجهٔ آن در جدول ۱ آمده است.

جدول ۱. نتایج بهترین ساختار شبکهٔ عصبی برای روش‌های نرمال‌سازی

روش‌های نرمال‌سازی داده‌ها	تعداد لایهٔ پنهان بهینه	تعداد نورون بهینهٔ لایهٔ پنهان	MSE	R
مینیمم- ماکزیمم	۱	۱۳	۰/۱۲	۰/۹۲
رتبه‌ای	۱	۳۰	۰/۲۳	۰/۹۷
نرمال استاندارد	۳	۱۰ ،۲۰ ،۱۵	۰/۳۱	۰/۸۲
داده‌های خام	۱	۲۶	۰/۴۷	۰/۶۷

1. Mean Square Error

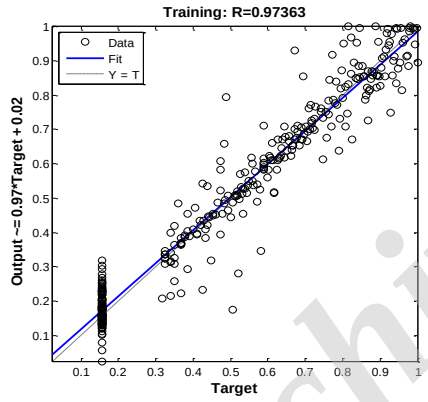
Archive of SID

www.SID.ir

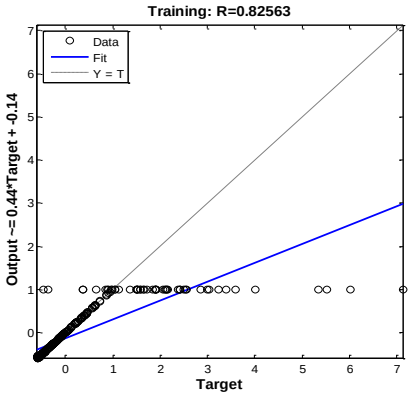
پس از انتخاب بهترین ساختار شبکه برای هر روش، تعمیم‌دهی شبکه‌ها با استفاده از کد مربوط به پیش‌بینی بررسی شد و در نهایت شبکه‌ای که بهترین تعمیم‌دهی را در مجموعهٔ آزمون و تست نشان داد به‌عنوان شبکهٔ مطلوب و روش به‌کار برده‌شده برای نرمال‌سازی به‌عنوان بهترین روش برای داده‌های بررسی‌شده انتخاب شد. در این پژوهش روش مینیمم‌ـ ماکزیمم با ساختار شبکهٔ سه‌لایه و تعداد ۱۳ نورون در لایهٔ پنهان با مقدار R= ۰/۹۲ و MSE =۰/۱۲ در مقایسه با دیگر روش‌ها به‌عنوان بهترین روش انتخاب شد. در حالی که عرفانیان و همکارانش (۱۳۹۲) با مقایسهٔ چهار روش پیش‌پردازش شامل رگرسیون بدون محدودیت داده‌ها، رگرسیون با اعمال محدودیت داده‌های ورودی، تست گاما و روش نرمال‌سازی مینیمم‌ـ ماکزیمم برای نرمال‌سازی داده‌های بارش و دمای متوسط ماهانه، بهترین نتایج را برای شبکهٔ عصبی با پیش‌پردازش داده‌های ورودی به روش رگرسیون بدون محدودیت ورودی به‌دست آوردند [۲۶]. گلکار و همکارانش (۱۳۸۸) نیز تابع لگاریتمی را به‌عنوان بهترین تابع نرمال‌ساز در بررسی کاربرد شبکهٔ عصبی مصنوعی در پیش‌بینی بارش منطقهٔ شیراز انتخاب کردند [۲۷]. در حالی که نصری و همکارانش (۱۳۸۹) به‌دلیل واریانس شدید مقادیر بارندگی و وجود تعداد زیاد روزهای بدون بارندگی (بارندگی صفر) از تبدیل نرمال استاندارد استفاده کردند [۱۹].

منحنی‌های پراکندگی مقادیر پیش‌بینی‌شدهٔ مدل در مقابل داده‌های واقعی برای هریک از روش‌های نرمال‌سازی در شکل‌های ۲ تا ۵ ارائه شده است. در این منحنی‌ها محور افقی نشان‌دهندهٔ مقدار واقعی و محور عمودی نشان‌دهندهٔ خروجی شبکه است. اختلاف بین محدودهٔ مقادیر در محورهای افقی به‌دلیل نرمال‌سازی به روش‌های مختلف است که هریک از این روش‌ها داده‌های خام را بسته به فرمول خود به محدوده‌های متفاوتی تبدیل می‌کنند.

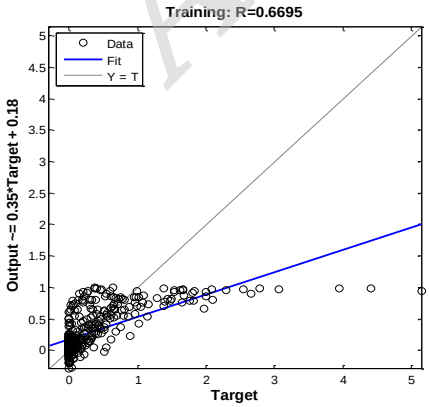
شکل ۲. مقایسهٔ مقادیر واقعی بارش ماهانه با میزان بارش پیش‌بینی‌شدهٔ شبکهٔ عصبی با روش مینیمم‌ـ ماکزیمم (با یک لایهٔ پنهان)



شکل ۳. مقایسهٔ مقادیر واقعی بارش ماهانه با میزان بارش پیش‌بینی‌شدهٔ شبکهٔ عصبی با روش نرمال‌سازی رتبه‌ای (با یک لایهٔ پنهان)



شکل ۴. مقایسهٔ مقادیر واقعی بارش ماهانه با میزان بارش پیش‌بینی‌شدهٔ شبکهٔ عصبی با روش نرمال استاندارد (با سه لایهٔ پنهان)



شکل ۵. مقایسهٔ مقادیر واقعی بارش ماهانه با میزان بارش پیش‌بینی‌شدهٔ شبکهٔ عصبی داده‌های خام و غیرنرمال (با یک لایهٔ پنهان)

Archive of SID
www.SID.ir

نتایج آنالیز حساسیت مدل با روش و ساختار شبکهٔ بهینه یعنی روش مینیمم- ماکزیمم و با ساختار شبکهٔ ۱-۱۳-۵ (دارای پنج ورودی شامل بیشترین رطوبت، کمترین رطوبت، بیشترین دما، کمترین دما و بارش ماه قبل، ۱۳ نورون در لایهٔ پنهان و بارش به‌عنوان خروجی شبکه) در جدول ۲ آمده است. با توجه به نتایج به‌دست‌آمده حذف پارامتر بیشترین رطوبت با مقدار MSE=۰/۲۸ بیشترین تأثیر و بارش ماه قبل با MSE=۰/۱۳ کمترین تأثیر را در پیش‌بینی بارش داشت.

همچنین نتایج مختص به تعداد ورودی‌های مختلف به شبکهٔ عصبی مصنوعی به شرح جدول ۳ است که نشان می‌دهد حتی با تعداد ۲ ورودی که شامل کمترین دما و کمترین رطوبت است می‌توان از نتیجهٔ خوبی به‌منظور پیش‌بینی بارش بهره‌مند شد.

جدول ۲. نتایج آنالیز حساسیت روش مینیمم- ماکزیمم

پارامتر حذف‌شده	MSE	تعداد ورودی‌ها
-	۰/۱۲	۵
بیشترین رطوبت	۰/۲۸	۴
بیشترین دما	۰/۲۷	۴
کمترین رطوبت	۰/۲۶	۴
کمترین دما	۰/۱۹	۴
بارش ماه قبل	۰/۱۳	۴

جدول ۳. نتایج انتخاب تعداد ورودی‌های مختلف به شبکه

پارامترهای ورودی	تعداد ورودی	MSE
کمترین دما، بیشترین دما، کمترین رطوبت، بیشترین رطوبت، بارش ماه قبل	۵	۰/۱۲
کمترین دما، بیشترین دما، کمترین رطوبت، بیشترین رطوبت	۴	۰/۱۳
کمترین دما، بیشترین دما، بارش ماه قبل	۳	۰/۱۴
کمترین دما، کمترین رطوبت	۲	۰/۱۳

نتیجه‌گیری

در این پژوهش ابتدا داده‌های گم‌شده و پرت مشخص شدند و سپس با استفاده از سه روش نرمال‌سازی شامل روش مینیمم- ماکزیمم، روش نرمال‌سازی رتبه‌ای و روش آمارهٔ نرمال استاندارد (z-score) نرمال شدند و از شبکهٔ عصبی MLP با تعداد لایه‌های پنهان و نورون‌های مختلف و الگوریتم یادگیری لونبرگ- مارکوارت برای تخمین بارش ماهانهٔ شهرستان آباده استفاده شد. از مقایسهٔ بهترین ساختارهای هر روش با یکدیگر، روش مینیمم- ماکزیمم با ساختار شبکهٔ سه لایه و تعداد ۱۳ نورون در لایهٔ پنهان با مقدار R= ۰/۹۲ و MSE =۰/۱۲ در مقایسه با دیگر روش‌ها به‌عنوان بهترین روش انتخاب شد. همچنین با مقایسهٔ عملکرد شبکه در استفاده از داده‌های خام و داده‌های پیش‌پردازش‌شده مشاهده شد که پیش‌پردازش داده‌ها تا حد زیادی سبب بهبود عملکرد شبکه شد. نتایج آنالیز حساسیت انجام‌شده نیز نشان داد مدل به حذف پارامتر بیشترین رطوبت بیشتر از سایر پارامترها حساسیت نشان داد. پس از آن نیز دمای حداکثر بیشترین تأثیر را بر پیش‌بینی بارش داشت. همچنین مقایسهٔ عملکرد شبکه با تعداد ورودی‌های مختلف نشان داد شبکه با داشتن دو ورودی کمترین دما و کمترین رطوبت با مقدار MSE =۰/۱۳ در مواقعی که کمبود داده وجود دارد نسبت به تعداد ۵ ورودی به نتیجهٔ مطلوبی رسید.

با توجه به نتایج یادشده می‌توان نتیجه گرفت که انجام پیش‌پردازش، به‌خصوص در پژوهش‌های مشابه می‌تواند سبب بهبود کارایی و دقت پیش‌بینی نتایج شبکه‌های عصبی شود. همچنین یکی دیگر از مزایای انجام پیش‌پردازش داده‌های ورودی صرفه‌جویی در زمان انجام تحلیل‌های شبکهٔ عصبی است، که دربارهٔ ساختن و مرحلهٔ آموزش شبکه برای داده‌های بدون پیش‌پردازش بسیار وقت‌گیر است و حتّی در بعضی از کامپیوترهای موجود قابل اجرا نیست.

Archive of SID

www.SID.ir

منابع

[1].Lookzadeh S. Evaluation of some methods for reconstruction of rainfall data in Alborz region, MSc, thesis, Tehran University,2004; P. 96. [In Persian].

[2].Kohzadi N, Boyd M, Kaastra I, Kermanshahi B, Scuse D. Neural networks for forecasting: an introduction. Canadian Journal of Agricultural Economics. 1995; 43: 463-474.

[3].Naghdi R, Shayannejhad M, Sadatinejad S.J. Comparison of different methods for estimating of monthly discharge missing data in Grand Karoon River Basin. Journal of Watershed Management Research. 2010; 1(1): 59-73. [In Persian].

[4].Mohammadi Takami M. The methods of data processing and pattern recognition. Faculty of electrical engineering. Khaje Nasireddin Toosi University. 2005.[In Persian].

[5].Nazeri Tahrudi M, Khalili K, Abbaszade Afshar M, Nazeri Tahrudi Z. Compared to the normal mechanism becomes the normal monthly rainfall data from different regions of Iran. Journal of Water and soil. 2014; 28 (2): 365-372. [In Persian].

[6].Adab H, Fallah Ghalhari Gh, Mirzabayati R. Evaluation of interpolation methods of Kriging and linear regression based on the DEM to mapping annual rainfall in Khorasan Razavi province. Geomatics Conference. Tehran. National Cartographic Center, Iran. 2008.[In Persian].

[7].Hamidi R, Emamgholizade S. Stochastic modeling of Maroon River annual discharge using ARMA model. The first National Conference of Applied Research of Iran Water Resources. Tehran. 2009. [In Persian].

[8].Shafiei M, Ghahraman B, Ansari H, Sharifi M. B. Stochastic simulation of drought severity based on Palmer Index. Journal of Water and Irrigation Management. 2011; 1 (1).1-13. [In Persian].

[9].Saghafian B, Razmkhah H, Ghermezcheshmeh B. Investigation of regional variations in annual rainfall by geostatistical methods, case study: Fars Province. Journal of Water Resources Engineering.2011; 4 (9): 29-38. [In Persian].

[10]. Ahmadi F, Dinpajouh Y, Fakherifard A, Khalili K. Modeling of river discharge using time series linear models (case study: Barandoozchay River). The first national conference of Strategies to achieve sustainable development in agriculture, natural resources and the environment. Tehran. 2012.[In Persian].

[11]. Nikmanesh M, Taleb Bidokhti N. Comparison of wavelet theory and time series in modeling of monthly rainfall of Saadatshahr and Arsenjan in Fars Province. Journal of Natural Geography.2012; 5 (16): 1-10. [In Persian].

[12]. Nazeri Tahrudi M, Khalili K, Ahmadi F, Nazeri Tahrudi Z. Modeling of temperature using periodic ARMA model (case study: Kerman Synoptic Station). Conference of applied researches in science and engineering. Tehran. 2012.[In Persian].

[13]. Nazeri Tahrudi M, Ahmadi F, Khalili K, Nazeri Tahrudi Z. Application of SAMS software to modeling of future climate of Kordestan Province (case study: Sanandaj Synoptic Station). The first conference of semi-arid regions hydrology. Sanandaj. 2013.[In Persian].

[14]. Nazeri Tahrudi M, Khalili K, Nazeri Tahrudi Z, Shahnazi M. Evaluation of ARIMA and PARMA models in modeling and forecasting maximum wind speed (case study: Bandar Abbas Synoptic Station). National conference of applied researches in science and engineering. Takestan. 2013.[In Persian].

[15]. Azadi S, Sepaskhah A. R. Annual precipitation forecast for west, southwest, and south provinces of Iran using artificial neural networks. Theoretical and Applied Climatology. 2012; 109(1): 175.

[16]. Leahy P, Kiely G, Corcoran G. Structural optimisation and input selection of an artificial neural network for river level prediction. J Hydrol,. 2008; 355:192–201.

[17]. Tabari H, Marofi S, Sabziparvar A.A. Estimation of daily pan evaporation using artificial neural network and multivariate non-linear regression. Irrigation Science. 2010; 28: 399-406.

[18]. Kim IS, Son JS, Park CE, Kim IJ, Kim HH. An investigation into an intelligent system for predicting bead geometry in Arc welding process. Int. J. of Materials Processing Technology. 2005; 159: 113–118.

[19]. Nasri M, Modarrs R, Dastoorani MT. Validation of ANN model of rainfall- runoff relationship in Zaynderood Dam Watershed. Journal of Watershed Researches.2010; 88: 17-26. [In Persian].

[20]. Hagan MT, Menhaj M. Training feed-forward networks with the Marquardt algorithm. IEEE Transactions on Neural Networks. 1994; 5(6): 989-993.

Archive of SID

www.SID.ir

[21]. Aksoy S, Haralick M. Feature Normalization and Likelihood-based Similarity Measures for Image Retrieval. Intelligent Systems Laboratory, Department of Electrical Engineering, University of Washington, Seattle, WA 98195-2500, U.S.A. 2000.

[22]. Nawi NM, Atomi WH, Rehman MZ. The Effect of Data Pre-Processing on Optimized Training of Artificial Neural Networks. Procedia Technology. 2003; 11: 32 – 39.

[23]. Willmotte CJ. Some comments on the evaluation of model performance. American Meteorological Society,1982; 63:1309-1313.

[24]. Haliban AH, Darand M. Forecasting rainfall using ANN. Journal of Applied Researches of Geographic Sciences. 2012; 12 (26): 47-63. [In Persian].

[25]. Amiri MJ, Abedi-Koupai J, Eslamian SS, Mousavi SF, Hasheminejad H. Modeling Pb(II) adsorption from aqueous solution by ostrich bone ash using adaptive neural-based fuzzy inference system, J. Environ. Sci. Health, Part A 48. 2013; 543–558.

[26]. Erfanian M, Ansari H, Alizade A. Forecasting monthly rainfall and average temperature using remote link templates with ANN (case study: Mashhad Synoptic Station). Geographical Studies of Arid Regions.2013; 3 (11): 53-73.[In Persian].

[27]. Golkar F, Farahmand A, Farahmand F. Evaluation of ANN application in prediction of Shiraz rainfall. National conference of Water Crisis Management. Islamic Azad University of Marvdasht. 2009.[In Persian].

Archive of SID

www.SID.ir